

## CLASSES OF TWO-DIMENSIONAL LANGUAGES AND RECOGNIZABILITY CONDITIONS \*

MARCELLA ANSELMO<sup>1</sup> AND MARIA MADONIA<sup>2</sup>

**Abstract.** The paper deals with some classes of two-dimensional recognizable languages of “high complexity”, in a sense specified in the paper and motivated by some necessary conditions holding for recognizable and unambiguous languages. For such classes we can solve some open questions related to unambiguity, finite ambiguity and complementation. Then we reformulate a necessary condition for recognizability stated by Matz, introducing a new complexity function. We solve an open question proposed by Matz, showing that all the known necessary conditions for recognizability of a language and its complement are not sufficient. The proof relies on a family of languages defined by functions.

**Mathematics Subject Classification.** 68Q45, 68Q70.

### 1. INTRODUCTION

The increasing interest for pattern recognition and image processing has motivated, in the sixties, the research on languages of *pictures* or *two-dimensional languages*, and nowadays this is a research field of great interest. Since then, many approaches have been presented in the literature in order to generalize the well-founded theory of formal languages of strings, and various models of finite automata, grammars, logics and regular expressions have been proposed. In [13,14] the family REC of *recognizable picture languages* was defined as an equivalent of

---

*Keywords and phrases.* Two-dimensional languages, unambiguity, complement.

\* This work was partially supported by MIUR Project “Aspetti matematici e applicazioni emergenti degli automi e dei linguaggi formali” (2007), by ESF Project “AutoMathA” (2005-2010), by 60% Projects of University of Catania and University of Salerno (2008, 2009).

<sup>1</sup> Dipartimento di Informatica ed Applicazioni, Università di Salerno, 84084 Fisciano (SA), Italy; [anselmo@dia.unisa.it](mailto:anselmo@dia.unisa.it)

<sup>2</sup> Dipartimento di Matematica e Informatica, Università di Catania, Viale Andrea Doria 6/a, 95125 Catania, Italy; [madonia@dmi.unict.it](mailto:madonia@dmi.unict.it)

the class of recognizable (or regular) string languages. This definition takes as starting point a characterization of recognizable string languages in terms of local languages and projections (*cf.* [11]). A two-dimensional language is said *recognizable* when it is the alphabetic projection of a local language defined in terms of a finite set of  $2 \times 2$  pictures called tiles; the recognition is given by a so called *tiling system*.

REC family inherits several properties from the class of regular string languages. A crucial difference lies in the fact that REC family is not closed under complementation: there are languages in REC whose complement is not in REC [14]. It is then important to take into account also the class co-REC of languages whose complement is in REC. The strict inclusion  $\text{REC} \subset (\text{REC} \cup \text{co-REC})$  holds even in the unary case [22] and it fits the fact that the definition of recognizability by tiling systems is intrinsically non-deterministic. The notion of determinism on tiling systems have been recently discussed in [4,6,20].

The non-closure of REC under complementation motivated the introduction of *unambiguous two-dimensional languages*, whose family is denoted UREC [13]. Informally, a picture language belongs to UREC when it admits an unambiguous tiling system, that is if every picture has a unique pre-image in its corresponding local language. In [3], the proper inclusion of UREC in REC is proved; it holds true in the unary case too (see [1]). In other words there exist in REC inherently ambiguous languages. An open question is whether UREC is closed under complementation or not. Its answer depends on the following open problem, where  $\bar{L}$  denotes the complement of language  $L$ .

*Question 1.* Does  $L \in \text{REC}$  and  $\bar{L} \notin \text{REC}$  imply that  $L \notin \text{UREC}$ ?

Question 1 was firstly stated in [25]. The converse is actually an open question too: Does  $L \in \text{REC} \setminus \text{UREC}$  imply that  $\bar{L} \notin \text{REC}$ ? Note that positive answers to both Question 1 and the converse mean that  $\text{UREC} = \text{REC} \cap \text{co-REC}$  and that UREC is the largest subset of REC closed under complementation. Also note that such question is related to some difficult problems on complexity classes [7].

All the inherently ambiguous languages known in the literature are *infinitely ambiguous*, in the sense that it is not possible to recognize them by a tiling system, in such a way that each picture has a fixed number of pre-images at most (see Sect. 2 for more details). The question whether this is always the case or not is open. Let us state it as follows.

*Question 2.* Does there exist a language  $L \in \text{REC} \setminus \text{UREC}$  such that  $L$  is finitely ambiguous?

In this paper we will answer Questions 1 and 2 in some particular cases, where languages involved have “high complexity”, as specified in the following. We will introduce a class  $\text{HP} \subseteq \text{co-REC} \setminus \text{REC}$  of not-recognizable languages and a class  $\text{HK} \subseteq \text{REC} \setminus \text{UREC}$  of ambiguous recognizable languages, whose languages are “hard” with respect to some complexity functions. We will show that:

1. If  $L \in \text{REC}$  and  $\bar{L} \in \text{HP}$  then  $L \notin \text{UREC}$ .
2. If  $L \in \text{REC}$  and  $L \in \text{HK}$  then  $L$  is infinitely ambiguous.

Let us emphasize that it is not known whether the inclusions  $\text{HP} \subseteq \text{co-REC} \setminus \text{REC}$  and  $\text{HK} \subseteq \text{REC} \setminus \text{UREC}$  are strict or not. No example (nor a candidate) exists

showing the inclusions are strict. Hence in the case  $HP = \text{co-REC} \setminus \text{REC}$  and/or  $HK = \text{REC} \setminus \text{UREC}$ , our results would be an answer to Questions 1 and/or 2, in their general setting.

The introduction of classes HP and HK is motivated by some necessary conditions for languages in REC and in UREC, respectively, stating that: if  $L \in \text{REC}$  then the size of some permutation matrices associated to  $L$  cannot grow so quickly; and if  $L \in \text{UREC}$  then the rank of the same matrices associated to  $L$  cannot grow so quickly. Here “HP” stands for “High Permutation” and “HK” stands for “High rank”. In the literature, all examples of languages that witness the strict inclusions  $\text{UREC} \subset \text{REC} \subset (\text{REC} \cup \text{co-REC})$  have been provided applying the necessary conditions we have just mentioned.

The main difficulty in this framework is that there are no characterizations of REC and UREC, that could be easily and fruitfully applied. A question that naturally arises is whether these conditions are also sufficient, so that they could characterize the classes. In this paper we will show that the considered conditions, that are necessary for the recognizability of a language or its complement, unfortunately are not sufficient.

Let us give some more details on the ideas on which the mentioned necessary conditions are based, since our results will be basically related to them.

In [21] Matz isolated a technique to obtain necessary conditions for the belonging of a picture language to REC. It consists in reducing the problem from two dimensions to one dimension and applying some known lower bounds on string automata. More precisely, for any recognizable picture language  $L$  and integer  $m$ , he considered the string language  $L(m)$  of all pictures in  $L$  of fixed height  $m$ . He proved that, if  $L \in \text{REC}$ , then it is possible to associate to any tiling system recognizing  $L$  a family  $\{A_m\}$ , where, for some constant  $c$ , each  $A_m$  is an automaton accepting  $L(m)$  with  $c^m$  states at most. Finally, he used the extended fooling set technique as a lower bound on the size of a string automaton [8,18].

Later on, in [3], Matz’s technique was used together with a lower bound on the size of unambiguous string automata. This provided a necessary condition for unambiguous picture languages, based on the growth of some parameters of the Hankel matrices of the string languages  $L(m)$ . Recently, in [15,16], the idea of finding necessary conditions for picture languages, using Matz’s technique, has been considered again introducing some complexity functions of Hankel matrices of  $L(m)$ . It provided a necessary condition for belonging to  $\text{REC} \cup \text{co-REC}$ , by rephrasing Cervelle’s one (see [10]), and a necessary condition for belonging to REC, using the fooling set technique [17] (instead of the extended fooling set technique) as a lower bound on the size of a string automaton.

Moreover, in 1998, Matz wondered whether the necessary condition for REC in [21] is also sufficient, in order to have a handy characterization of REC. He proposed as a counter-example the language of pictures over  $\{a, b\}$  with same number of  $a$ ’s as  $b$ ’s, that satisfies his condition. But then Reinhardt [26] proved that the language is recognizable. Since then, other candidate languages have been proposed as counter-examples, but they always failed. Here we negatively answer

the open question, after rephrasing Matz's condition in the framework of Hankel matrices.

The analogous question regarding other necessary conditions for REC and for  $\text{REC} \cup \text{co-REC}$  can be considered. We prove that all these conditions are not sufficient. The counter-examples are given by a family of languages defined by functions. Recently another counter-example showing that Matz's condition is not sufficient was independently found (see [9]) by a completely different technique. This technique has provided a family of binary languages, whereas our counter-example is a family of unary languages.

The paper is organized as follows. After giving the basic definitions and results on two-dimensional languages in Section 2, in Section 3 we recall the above mentioned necessary conditions for two-dimensional languages and introduce the classes HP and HK. Section 4 contains the main results concerning Questions 1 and 2. In Section 5 we introduce a new complexity function of picture languages that allows us to reformulate Matz's condition in terms of the growth of this function. At last in Section 6 we show that the considered recognizability conditions are not sufficient.

Some results of this paper can be found in [2].

## 2. PRELIMINARIES

In this section we recall some definitions about two-dimensional recognizable languages. More details can be mainly found in [14].

A *two-dimensional string* (or a *picture*) over a finite alphabet  $\Sigma$  is a two-dimensional rectangular array of elements of  $\Sigma$ . The set of all pictures over  $\Sigma$  is denoted by  $\Sigma^{**}$  and a *two-dimensional language* over  $\Sigma$  is a subset of  $\Sigma^{**}$ .

Given a picture  $p \in \Sigma^{**}$ , let  $p_{(i,j)}$  denote the symbol in  $p$  with coordinates  $(i, j)$ ,  $\ell_1(p)$ , the number of rows and  $\ell_2(p)$  the number of columns; the pair  $(\ell_1(p), \ell_2(p))$  is the *size* of  $p$ . Note that when a one-letter alphabet is concerned, a picture  $p$  is totally defined by its size  $(m, n)$ , and we will write  $p = (m, n)$ . Remark that the set  $\Sigma^{**}$  includes also all the empty pictures, *i.e.* all pictures of size  $(m, 0)$  and of size  $(0, n)$ , for all  $m, n \geq 0$ , that we denote by  $\lambda_{m,0}$  and  $\lambda_{0,n}$ , respectively. The set of all pictures over  $\Sigma$  of size  $(m, n)$  is denoted by  $\Sigma^{m,n}$ , while  $\Sigma^{m,*}$  denotes the set of all pictures over  $\Sigma$  with  $m$  rows. It will be needed to identify the symbols on the boundary of a given picture: for any picture  $p$  of size  $(m, n)$ , we consider the *bordered picture*  $\hat{p}$  of size  $(m+2, n+2)$  obtained by surrounding  $p$  with a special *boundary symbol*  $\# \notin \Sigma$ .

A *tile* is a picture of size  $(2, 2)$  and  $B_{2,2}(p)$  is the set of all sub-blocks of size  $(2, 2)$  of a picture  $p$ . Given an alphabet  $\Gamma$ , a two-dimensional language  $L \subseteq \Gamma^{**}$  is *local* if there exists a set  $\Theta$  of tiles over  $\Gamma \cup \{\#\}$  such that  $L = \{p \in \Gamma^{**} \mid B_{2,2}(\hat{p}) \subseteq \Theta\}$  and we will write  $L = L(\Theta)$ .

A *tiling system* is a quadruple  $(\Sigma, \Gamma, \Theta, \pi)$  where  $\Sigma$  and  $\Gamma$  are finite alphabets,  $\Theta$  is a finite set of tiles over  $\Gamma \cup \{\#\}$  and  $\pi : \Gamma \rightarrow \Sigma$  is a projection. A two-dimensional language  $L \subseteq \Sigma^{**}$  is *tiling recognizable* if there exists a tiling system  $(\Sigma, \Gamma, \Theta, \pi)$

such that  $L = \pi(L(\Theta))$  (extending  $\pi$  in the usual way). For any  $p \in L$ , a local picture  $p' \in L(\Theta)$ , such that  $p = \pi(p')$ , is called a *pre-image* of  $p$ . We denote by REC the family of all *tiling recognizable* picture languages.

The family REC is closed with respect to many operations. The *column concatenation* of  $p$  and  $q$  (denoted by  $p \Phi q$ ) and the *row concatenation* of  $p$  and  $q$  (denoted by  $p \Theta q$ ) are partial operations, defined only if  $\ell_1(p) = \ell_1(q)$  and if  $\ell_2(p) = \ell_2(q)$ , respectively, and are given by:

$$p \Phi q = \begin{array}{|c|c|} \hline p & q \\ \hline \end{array} \qquad p \Theta q = \begin{array}{|c|} \hline p \\ \hline q \\ \hline \end{array}$$

As done in the string language theory, these definitions of picture concatenations can be extended to define two-dimensional languages concatenations. If  $L_1, L_2$  are picture languages over an alphabet  $\Sigma$ , the *column concatenation* and the *row concatenation* of  $L_1$  and  $L_2$  are defined respectively by  $L_1 \Phi L_2 = \{p \Phi q \mid p \in L_1, q \in L_2\}$  and  $L_1 \Theta L_2 = \{p \Theta q \mid p \in L_1, q \in L_2\}$ . Furthermore, by iterating the concatenation operations, we obtain the column and row *closure* or *star*. More precisely: the *column closure* of  $L$  (denoted by  $L^{*\Phi}$ ) and the *row closure* of  $L$  (denoted by  $L^{*\Theta}$ ) are defined as  $L^{*\Phi} = \bigcup_i L^{i\Phi}$  and  $L^{*\Theta} = \bigcup_i L^{i\Theta}$  where  $L^{0\Phi} = \{\lambda_{m,0} \mid m \geq 0\}$ ,  $L^{n\Phi} = L^{(n-1)\Phi} \Phi L$  and  $L^{0\Theta} = \{\lambda_{0,m} \mid m \geq 0\}$ ,  $L^{n\Theta} = L^{(n-1)\Theta} \Theta L$ .

REC family is closed under row and column concatenation and their closures, under union, intersection and under rotation (see [14] for all the proofs).

Let us give some examples.

**Example 2.1.** Let  $L_{fc=lc}$  be the language of pictures over  $\Sigma = \{a, b\}$ , with more than one column, whose first column is equal to the last one. Language  $L_{fc=lc} \in$  REC. Informally we can define a local language where information about first column symbols of a picture  $p$  is brought along horizontal direction, by means of subscripts, to match the last column of  $p$ . Tiles are defined to have always the same subscripts within a row; moreover in left and right border tiles, subscripts and main symbols should match. Below it is an example of a picture  $p \in L_{fc=lc}$  together with a pre-image  $p'$  of  $p$ .

$$p = \begin{array}{|c|c|c|c|c|} \hline b & b & a & b & b \\ \hline a & a & b & a & a \\ \hline b & a & a & a & b \\ \hline a & b & b & b & a \\ \hline \end{array} \qquad p' = \begin{array}{|c|c|c|c|c|} \hline b_b & b_b & a_b & b_b & b_b \\ \hline a_a & a_a & b_a & a_a & a_a \\ \hline b_b & a_b & a_b & a_b & b_b \\ \hline a_a & b_a & b_a & b_a & a_a \\ \hline \end{array}$$

Let  $L_{fc=c'}$  be the language of pictures such that the first column is equal to some  $i$ -th column,  $i \neq 1$ . Note that  $L_{fc=c'} = L_{fc=lc} \Phi \Sigma^{**}$  and thus  $L_{fc=c'} \in$  REC. Similarly we can show that the languages  $L_{c'=lc} = \Sigma^{**} \Theta L_{fc=lc}$ , and  $L_{c=c'} = \Sigma^{**} \Phi L_{fc=lc} \Phi \Sigma^{**}$  are in REC.

**Example 2.2.** Consider the language *CORNERS* of all pictures  $p$  over  $\Sigma = \{a, b\}$  such that whenever  $p_{(i,j)} = p_{(i',j)} = p_{(i,j')} = b$  then also  $p_{(i',j')} = b$ . Intuitively, whenever three corners of a rectangle inside a picture all carry symbol  $b$ , then also

the fourth one does. In [21], it is shown that  $CORNERS \notin REC$ . Consider now, the language  $L = \overline{CORNERS}$ . We have  $L \in REC$ ; indeed, we can set

$$L_1 = \Sigma^{**} \ominus (\Sigma^{**} \oplus (\boxed{b} \ominus \Sigma^{**} \ominus \boxed{b}) \oplus \Sigma^{**} \oplus (\boxed{b} \ominus \Sigma^{**} \ominus \boxed{a}) \oplus \Sigma^{**}) \ominus \Sigma^{**},$$

and then  $L$  is equal to the union of  $L_1$  with the languages obtained by its  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  rotations.

A recognizable two-dimensional language  $L \subseteq \Sigma^{**}$  is *unambiguous* if and only if it admits an unambiguous tiling system  $\mathcal{T}$ ; a tiling system  $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$  is *unambiguous* for  $L$  if and only if any picture  $p \in L$  has an unique pre-image in the local language  $L(\Theta)$  (see [13]). The family of all unambiguous recognizable two-dimensional languages is denoted by UREC. In [3] it is proved that the inclusion of UREC in REC is strict and in [1] that this strict inclusion holds even if the alphabet is unary. Therefore in REC there exist languages that are inherently ambiguous.

Let us now recall the definitions of  $k$ -ambiguity, finite and infinite ambiguity given in [5] for languages in REC. Note that a similar definition of  $k$ -ambiguity is contained in [24]. A tiling system  $\mathcal{T} = (\Sigma, \Gamma, \Theta, \pi)$  recognizing  $L$  is said to be *k-ambiguous* if every picture  $p \in L$  has at most  $k$  pre-images. A recognizable language  $L$  is said *k-ambiguous* if  $k = \min\{s \mid \mathcal{T} \text{ is } s\text{-ambiguous tiling system and } \mathcal{T} \text{ recognizes } L\}$ . A language  $L$  is *finitely ambiguous* if it is  $k$ -ambiguous for some  $k$ , whereas a language  $L$  is *infinitely-ambiguous* if it is not finitely ambiguous.

### 3. CLASSES HP AND HK

In this section we introduce the definitions of the classes HP and HK of picture languages of high complexity (“HP” stands for “High Permutation”, and “HK” stands for “High rank”) motivated by some recognizability conditions we recall as well. These conditions are based on the idea of reducing the problem from two dimensions to one dimension and, then, using some known lower bounds on string languages.

Let  $L \subseteq \Sigma^{**}$  be a picture language. For any  $m \geq 1$ , we can consider the subset  $L(m) \subseteq L$  containing all pictures in  $L$  with exactly  $m$  rows. Note that the language  $L(m)$  can be viewed as a string language over the alphabet of the columns of height  $m$ .

Recall that a *nondeterministic automaton (NFA)* over an alphabet  $\Sigma$  is a quadruple  $\mathcal{A} = (Q, q_0, F, \delta)$  where  $Q$  is the finite set of states,  $\delta: Q \times \Sigma \rightarrow 2^Q$  is the transition function,  $q_0 \in Q$  is the initial state,  $F \subseteq Q$  is the set of final states. The transition function  $\delta$  can be extended to strings to  $\hat{\delta}$  in a standard way. The *language accepted by  $\mathcal{A}$*  is  $\mathcal{L}(\mathcal{A}) = \{w \in \Sigma^* \mid \hat{\delta}(q_0, w) \cap F \neq \emptyset\}$ .

Let us state the recognizability condition proved by Matz (Lem. 3.2). It is based on Lemma 3.1 (that relates the recognizability of a two-dimensional language with the complexity of some associated one-dimensional languages) together with the extended fooling set technique (that gives a lower bound on the complexity of a one-dimensional language). Remark that Lemmata 3.1 and 3.2 can be used to show that a language is not recognizable (as for language  $CORNERS$  in Ex. 2.2).

**Lemma 3.1** ([21]). *If  $L$  is in  $REC$  then it is possible to associate to any tiling system recognizing  $L$  a constant  $c$  and a family  $\{\mathcal{A}_m\}$  such that each  $\mathcal{A}_m$  is an automaton accepting  $L(m)$  with  $c^m$  states at most.*

**Lemma 3.2** ([21]). *Let  $L \subseteq \Sigma^{**}$ ,  $L \in REC$  and let  $\{P_m\}$  be a sequence such that, for any  $m$ ,  $P_m \subseteq \Sigma^{m,*} \times \Sigma^{m,*}$  and*

- (1) *for all  $(p, q) \in P_m$ ,  $p \Phi q \in L$ ;*
  - (2) *for all  $(p, q), (p', q') \in P_m$ , with  $(p, q) \neq (p', q')$ ,  $\{p \Phi q', p' \Phi q\} \not\subseteq L$ .*
- Then  $|P_m|$  is  $2^{O(m)}$ .*

Lemma 3.1 is the starting point for other necessary conditions we are going to deal with. They are formulated in terms of Hankel matrices. Recall that, for any string language  $L \subseteq \Sigma^*$ , the infinite boolean Hankel matrix associated to  $L$ , is  $M_L = \|a_{\alpha\beta}\|_{\alpha \in \Sigma^*, \beta \in \Sigma^*}$  where  $a_{\alpha\beta} = 1$  if and only if  $\alpha\beta \in L$  (see [19]). Observe that, when  $L$  is a regular language, the number of different rows of  $M_L$  is finite (Myhill-Nerode Theorem). A sub-matrix  $M_{(U,V)}$  of an Hankel matrix  $M_L$  is a matrix specified by a pair of languages  $(U, V)$ , with  $U, V \subseteq \Sigma^*$ , that is obtained by intersecting all rows and all columns of  $M_L$  that are indexed by the strings in  $U$  and  $V$ , respectively. Moreover, given a matrix  $M$ , we denote by  $Rank_Q(M)$ , the rank of  $M$  over the field of rational numbers  $Q$ . A *permutation matrix* is a boolean matrix that has exactly one 1 in each row and in each column.

The investigation of some parameters of the Hankel matrices associated to  $L(m)$ , when  $m$  grows, will provide a uniform setting for several recognizability conditions.

**Definition 3.1.** [15] Let  $L$  be a picture language.

- (i) The row complexity function  $R_L(m)$  gives the number of distinct rows of the matrix  $M_{L(m)}$ .
- (ii) The permutation complexity function  $P_L(m)$  gives the size of the maximal permutation matrix that is a sub-matrix of  $M_{L(m)}$ .
- (iii) The rank complexity function  $K_L(m)$  gives the rank of the matrix  $M_{L(m)}$ .

Remark that, obviously,  $K_L(m) \leq R_L(m)$  for any picture language  $L$  and integer  $m$ . The following theorem collects some necessary conditions for picture languages.

**Theorem 3.1.** *Let  $L \subseteq \Sigma^{**}$ .*

1. *If  $L \in REC \cup co-REC$  then there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $R_L(m) \leq 2^{c^m}$ .*
2. *If  $L \in REC$  then there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $P_L(m) \leq c^m$ .*
3. *If  $L \in UREC$  then there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $K_L(m) \leq c^m$ .*
4. *If  $L \in REC \setminus UREC$  and  $L$  is finitely ambiguous then there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $K_L(m) \leq c^m$ .*

*Proof.* Item 1 is essentially due to Cervelle [10]; it is rephrased in the matrix framework as in [15]. Item 2 is the analogous of Matz’s Lemma 3.2, but it is obtained applying the fooling set technique (instead of its extended version) and rephrasing the condition on Hankel matrices [15]. Item 3 is proved in [3]. Item 4 can be found in [24], for a bit different definition of  $k$ -ambiguity, but it holds

even for the definition presented in this paper. Indeed if  $L$  is  $k$ -ambiguous then there exists a constant  $c$  such that, for any  $m \geq 1$ , the automaton  $\mathcal{A}_m$ , defined in Lemma 3.1, that recognizes language  $L(m)$  with  $c^m$  states at most, is  $k$ -ambiguous. Then we can apply a lower bound on the number of states of  $k$ -ambiguous automata in [19] that guarantees that  $|Q_m| \geq \text{Rank}_Q(M_L)^{1/k} - 1$ , where  $Q_m$  denotes the set of states of  $\mathcal{A}_m$ . Therefore  $K_L(m)^{1/k} \leq d^m$ , for some constant  $d \in \mathbb{N}$ , and finally  $K_L(m) \leq (d^k)^m$ .  $\square$

Note that in [4,6], some subclasses of REC have been introduced and similar necessary conditions founded on  $R_L(m)$  have been proved for them. Now a question naturally rises: are all these necessary conditions also sufficient? We will come back to the question in Section 6.

We are now ready to introduce classes HP and HK.

**Definition 3.2.** *HP* is the class of all picture languages  $L \in \text{co-REC}$  for which there does not exist a constant  $c$  such that  $P_L(m) \leq c^m$ , for all  $m \geq 1$ .

*HK* is the class of all picture languages  $L \in \text{REC}$  for which there does not exist a constant  $c$  such that  $K_L(m) \leq c^m$ , for all  $m \geq 1$ .

From Theorem 3.1, if  $L \in \text{HP}$  then  $L \notin \text{REC}$  and if  $L \in \text{HK}$  then  $L \notin \text{UREC}$ . Let us now show some examples of languages in HP and in HK. We will use a technical result, proved in the following lemma, concerning the rank of some special boolean matrices. In the following, for any matrix  $A = \|a_{ij}\|$  with  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ,  $A_{ij}$  will denote the  $(i, j)$  minor of  $A$ .

**Lemma 3.3.** *Let  $A = \|a_{ij}\|$  be a boolean square matrix of size  $k$  such that, for any  $1 \leq i, j \leq k$ ,  $a_{ij} = 0$  if and only if  $i + j = k + 1$ . Then  $\text{Rank}_Q(A) = k$ .*

*Proof.* It suffices to prove that  $\det(A) \neq 0$ . Remark that  $A$  is a square matrix with 0 in all counter-diagonal positions and 1 elsewhere. Let us evaluate  $\det(A)$  along its first row:  $\det(A) = \sum_{i=1}^k (-1)^{1+i} a_{1i} \det(A_{1i}) = \det(A_{11}) + (-1) \det(A_{12}) + \dots + (-1)^k \det(A_{1k-1}) + 0 \det(A_{1k})$ .

Since, for any  $i = 2, \dots, k - 1$ , the matrix  $A_{1i}$  can be obtained from the matrix  $A_{1i-1}$  by swapping its  $(k - i + 1)$ -th row with its  $(k - i)$ -th one, we can say that  $\det(A) = (k - 1) \det(A_{11})$ . Therefore, in order to prove that  $\det(A) \neq 0$ , it suffices to show that  $\det(A)_{11} \neq 0$ . Note that  $A_{11}$  is a square matrix, of size  $k - 1$ , that has 0 in all the positions immediately above the counter-diagonal and 1 elsewhere. Let us denote by  $B^h$  the square matrix of size  $h$  of this form and let us show that, for any  $h$ ,  $B^h$  has a non-null determinant. The proof is by induction on  $h$ . The basis,  $h = 2$ , is obvious. Suppose that it is true for  $B^{h-1}$  and consider the matrix  $B^h = \|b_{ij}\|$ . If we evaluate  $\det(B^h)$  along its first column, we have  $\det(B^h) = \sum_{i=1}^h (-1)^{1+i} b_{i1} \det(B_{i1}^h)$ . Remark that the first  $(h - 2)$  terms of the sum are equal to 0 (every matrix  $B_{i1}^h$  has two identical rows, the last one and the second-last one, and therefore it has a null determinant) and the  $(h - 1)$ -th term is equal to 0 too (note that  $b_{(h-1)1} = 0$ ). So we have  $\det(B^h) = (-1)^{h+1} \det(B^{h-1})$  and, therefore, by inductive hypothesis,  $\det(B^h) \neq 0$ .  $\square$



Now, let us fix some notation: we denote by  $\varepsilon$  the empty string and, for  $\Sigma = \{a\}$  and  $n \in \mathbb{N}$ , by  $a^n$  the string over  $\Sigma^*$  of length  $n$ . Moreover, for  $n_1, n_2, \dots, n_m \in \mathbb{N}$ , we denote by  $lcm(n_1, n_2, \dots, n_m)$  the lowest common multiple of  $n_1, n_2, \dots, n_m$ .

**Example 3.1.** Consider, for any  $m \geq 0$ , the function  $f(m) = lcm(2^{m+1}, \dots, 2^{m+1})$  and the language  $L$  over the unary alphabet  $\Sigma = \{a\}$ ,  $L = \{(m, n) \mid n \text{ is not a multiple of } f(m)\}$ . In [22,23] it is shown that  $L \in \text{REC}$ .

Now, we will show that  $L \in \text{HK}$ . Indeed, for any  $m > 1$ , consider languages  $L(m)$  as defined above and the corresponding boolean matrix  $M = M_{L(m)}$ . Let us denote by  $v$  the picture over the alphabet  $\Sigma$  with  $m$  rows and one column and consider the set  $S$  of the  $f(m)$  rows of  $M$  indexed by  $v, v^2, \dots, v^{f(m)}$ . They are all distinct (for every  $i, j$  with  $1 \leq i < j \leq f(m)$ , the rows indexed by  $v^i$  and  $v^j$  differ in the position corresponding to the column indexed by  $v^{f(m)-i}$ ) and, moreover, any other row in  $M$  is equal to one of the rows in  $S$ . So  $R_L(m) = f(m)$ . Consider now, in  $M$ , the finite sub-matrix  $M_c$  composed by the  $f(m)$  rows indexed by  $v, v^2, \dots, v^{f(m)}$ , in this order, and the  $f(m)$  columns indexed by  $\varepsilon, v, v^2, \dots, v^{f(m)-1}$ , in this order, as in the following figure.

	$\varepsilon$	$v$	$v^2$	$\dots$	$v^{f(m)-3}$	$v^{f(m)-2}$	$v^{f(m)-1}$
$v$	1	1	1	$\dots$	1	1	0
$v^2$	1	1	1	$\dots$	1	0	1
$v^3$	1	1	1	$\dots$	0	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v^{f(m)-2}$	1	1	0	$\dots$	1	1	1
$v^{f(m)-1}$	1	0	1	$\dots$	1	1	1
$v^{f(m)}$	0	1	1	$\dots$	1	1	1

Then, the  $i$ -th row of  $M_c$ , for  $i = 1, 2, \dots, f(m)$ , carries symbol 1 in all its entries except the  $f(m) + 1 - i$  position that carries symbol 0. Matrix  $M_c$  satisfies the hypothesis of Lemma 3.3, and, therefore,  $Rank_Q(M_c) = f(m)$ . But  $f(m) = Rank_Q(M_c) \leq K_L(m) \leq R_L(m) = f(m)$ , so we have  $K_L(m) = f(m)$ . Since  $f(m) = 2^{\theta(2^m)}$  (see [21,22]), then  $K_L(m)$  cannot be bounded by  $c^m$ , where  $c$  is a constant, and therefore  $L \in \text{HK}$ .

At last, it is easy to see that, for any  $m > 1$ ,  $P_L(m) = 2$ .

Consider now the language  $\bar{L}$  and, for any  $m > 1$ , languages  $\bar{L}(m)$ . Remark that the Hankel matrix of the complement of a language can be simply obtained by exchanging 0 and 1 occurrences in the Hankel matrix for the language. Then the finite sub-matrix of  $M_{\bar{L}(m)}$ , with rows and columns indexes as in  $M_c$ , is a square matrix of size  $f(m)$  with 1 in all counter-diagonal positions and 0 elsewhere. We have that, for any  $m$ ,  $P_{\bar{L}}(m) = R_{\bar{L}}(m) = f(m)$  and, therefore,  $\bar{L} \in \text{HP}$ . Furthermore, it is easy to show that  $K_{\bar{L}}(m) = f(m)$  too.

**Example 3.2.** Let *CORNERS* be the language defined in Example 2.2. We are going to show that *CORNERS*  $\in$  HP, following the proof that *CORNERS*  $\notin$  REC in [21].

Consider for any  $n \geq 1$  a partition  $\mathcal{P}$  of  $\{1, 2, \dots, 2n\}$  into two-element sets and fix a bijection  $\alpha_{\mathcal{P}} : \mathcal{P} \rightarrow \{1, 2, \dots, n\}$ . Then define picture  $P_{\mathcal{P}}$  over  $\{a, b\}$  as the picture of size  $(2n, n)$  such that the position  $(i, j)$  in  $P_{\mathcal{P}}$  carries symbol  $b$  if and only if  $j = \alpha_{\mathcal{P}}(\{i, i'\})$  and  $\{i, i'\} \in \mathcal{P}$ . As an example let  $n = 3$ ,  $\mathcal{P} = \{(1, 2), (3, 4), (5, 6)\}$  and  $\mathcal{P}' = \{(1, 3), (2, 4), (5, 6)\}$ ; then fix  $\alpha_{\mathcal{P}}((1, 2)) = 1$ ,  $\alpha_{\mathcal{P}}((3, 4)) = 2$ , and  $\alpha_{\mathcal{P}}((5, 6)) = 3$ ;  $\alpha_{\mathcal{P}'}((1, 3)) = 1$ ,  $\alpha_{\mathcal{P}'}((2, 4)) = 2$ , and  $\alpha_{\mathcal{P}'}((5, 6)) = 3$ . Pictures  $P_{\mathcal{P}}$  and  $P_{\mathcal{P}'}$  are as follows:

$$P_{\mathcal{P}} = \begin{array}{|c|c|c|} \hline b & a & a \\ \hline b & a & a \\ \hline a & b & a \\ \hline a & b & a \\ \hline a & a & b \\ \hline a & a & b \\ \hline \end{array} \quad P_{\mathcal{P}'} = \begin{array}{|c|c|c|} \hline b & a & a \\ \hline a & b & a \\ \hline b & a & a \\ \hline a & b & a \\ \hline a & a & b \\ \hline a & a & b \\ \hline \end{array}$$

Let  $M_{L(2n)}$  be the Hankel matrix of the language  $L(2n)$  of pictures in *CORNERS* of fixed height  $2n$ , and  $M_{(U,V)}$  its sub-matrix specified by the pair of languages  $(U, V)$  with  $U = V = \{P_{\mathcal{P}} \mid \mathcal{P} \text{ is a partition of } \{1, 2, \dots, 2n\} \text{ into two-element sets}\}$ . We have that  $M_{(U,V)}$  is a permutation matrix. Indeed the entry  $(P_{\mathcal{P}}, P_{\mathcal{P}'})$  of  $M_{(U,V)}$  is 1 if and only if  $\mathcal{P} = \mathcal{P}'$ .

Furthermore the size of matrix  $M_{(U,V)}$  is equal to the number  $A_n$  of partitions of  $\{1, 2, \dots, 2n\}$  into two-element sets. And it can be shown that  $A_n \geq n!$  and then there does not exist  $c \in \mathbb{N}$  such that  $A_n \leq c^n$ . Hence *CORNERS*  $\in$  HP.

Let us mention that another language in HK is  $L_{c=c'}$  as introduced in Example 2.1 (see [4]), while its complement is in HP (see [15]).

#### 4. SOME RESULTS ON CLASSES HP AND HK

In this section we answer Questions 1 and 2 as stated in the Introduction, in the case the involved languages belong to classes HP and HK introduced in Section 3. Firstly let us compare the values of the complexity functions  $R_L(m)$ ,  $P_L(m)$  and  $K_L(m)$  introduced in Section 3, for a language  $L$  and its complement, in the case  $L$  is in  $\text{REC} \cup \text{co-REC}$  (and therefore functions  $R_L(m)$ ,  $P_L(m)$  and  $K_L(m)$  have finite values).

**Proposition 4.1.** *Let  $L \in \text{REC} \cup \text{co-REC}$ .*

1.  $R_{\overline{L}}(m) = R_L(m)$ .
2.  $P_L(m) + P_{\overline{L}}(m) \leq R_L(m) + 2$  and the bound is tight.
3.  $K_L(m) + K_{\overline{L}}(m) \leq 2R_L(m)$  and the bound is tight.
4.  $K_{\overline{L}}(m) \geq P_L(m)$  and the bound is tight.

*Proof.*

- (1) The Hankel matrices for  $\bar{L}$  can be obtained by exchanging entries 0 with entries 1 in the Hankel matrices for  $L$ .
- (2) Let  $m \geq 1$ ,  $M_{(U,V)}$  be a permutation matrix of maximal size that is a sub-matrix of the Hankel matrix  $M_{L(m)}$  for  $L(m)$ , and let  $M'_{(U',V')}$  be a permutation matrix of maximal size that is a sub-matrix of the Hankel matrix  $M_{\bar{L}(m)}$  for the complement of  $L(m)$ . We claim that  $|U \cap U'|, |V \cap V'| \leq 2$ . In other words the sub-matrices of  $M_{L(m)}$  specified by  $(U, V)$  and  $(U', V')$ , respectively, cannot overlap on a square matrix of size greater than 2. Consider indeed a column of  $M_{L(m)}$  indexed by a string in  $|V \cap V'|$ . The set of entries on such column indexed by strings in  $U$  are all 0's except for one 1 and then they cannot share more than two elements (a 0 and a 1) with the set of entries indexed by strings in  $U'$  (that are all 1's except for one 0).

The bound is tight for language  $L$  in Example 3.1:  $P_L(m) = 2$  and  $P_{\bar{L}}(m) = R_L(m) = f(m)$  and, therefore,  $P_L(m) + P_{\bar{L}}(m) = R_L(m) + 2$ .

- (3) The inequality follows from Item 1 and from the remark that  $K_L(m) \leq R_L(m)$ . The bound is tight for language  $L$  in Example 3.1:  $K_L(m) = K_{\bar{L}}(m) = R_L(m) = f(m)$  and, therefore,  $K_L(m) + K_{\bar{L}}(m) = 2R_L(m)$ .
- (4) Let  $P$  be a maximal permutation matrix of  $M_{L(m)}$ .  $P$  is a boolean square matrix of size  $P_L(m)$  and we can assume, without loss of generality, that  $P$  has 1 in all counter-diagonal positions and 0 elsewhere. Now, consider  $M_{\bar{L}(m)}$  and its sub-matrix of size  $P_L(m)$ , say  $P'$ , that corresponds to the permutation matrix  $P$  of  $M_{L(m)}$ . Remark that  $P'$  is a square matrix of size  $P_L(m)$  with 0 in all counter-diagonal positions and 1 elsewhere. Therefore, the matrix  $P'$  satisfies the hypothesis of Lemma 3.3 and we have  $Rank_Q(P') = P_L(m)$  that implies  $K_{\bar{L}}(m) \geq P_L(m)$ . The bound is tight for language  $\bar{L}$ , that is the complement of language  $L$  in Example 3.1. □

From Item 4 of Proposition 4.1, this result immediately follows.

**Corollary 4.1.** *If  $L \in HP$  then  $\bar{L} \in HK$ .*

The following proposition is a positive answer to Question 1 (see the introduction) in the case where  $\bar{L} \notin REC$  since  $\bar{L} \in HP$ . Recall that if a language of co-REC is in HP then it is necessarily not in REC; and that we do not know at present whether this is also a sufficient condition for languages in co-REC. Note that, if this condition in co-REC were also sufficient, then  $HP = co-REC \setminus REC$  and vice versa. In the case  $HP = co-REC \setminus REC$ , Proposition 4.2 would be a positive answer to Question 1 in its general setting.

**Proposition 4.2.** *If  $L \in REC$  and  $\bar{L} \in HP$  then  $L \notin UREC$ .*

*Proof.* If  $\bar{L} \in HP$  then, from Corollary 4.1,  $L \in HK$  and therefore, from the definition of HK and Theorem 3.1,  $L \notin UREC$ . □

The following proposition is a negative answer to Question 2 (see the introduction) in the case where  $L \notin \text{UREC}$  since  $L \in \text{HK}$ . Recall that if a language of  $\text{REC}$  is in  $\text{HK}$  then it is necessarily inherently ambiguous; and that we do not know at present whether this is also a sufficient condition for languages in  $\text{REC}$ . Note that, if this condition in  $\text{REC}$  were also sufficient, then  $\text{HK} = \text{REC} \setminus \text{UREC}$  and vice versa. In the case  $\text{HK} = \text{REC} \setminus \text{UREC}$ , Proposition 4.3 would be a negative answer to Question 2 in its general setting.

**Proposition 4.3.** *Any language  $L \in \text{REC} \setminus \text{UREC}$  such that  $L \in \text{HK}$  is infinitely ambiguous.*

*Proof.* The proof follows from Item 4 in Theorem 3.1.  $\square$

As an application, consider the following example.

**Example 4.1.** Consider the language *CORNERS* defined in Example 2.2. In Example 3.2 we showed that  $\text{CORNERS} \in \text{HP}$  and, applying Corollary 4.1, we have that  $\overline{\text{CORNERS}} \in \text{HK}$ . Moreover, since  $\overline{\text{CORNERS}} \in \text{REC}$  (see Ex. 2.2), applying Proposition 4.2, we have that  $\overline{\text{CORNERS}} \notin \text{UREC}$ . This implies, from Proposition 4.3, that  $\overline{\text{CORNERS}}$  is infinitely ambiguous.

## 5. ANOTHER RECOGNIZABILITY CONDITION FOR PICTURE LANGUAGES

A first necessary condition for recognizability of picture languages was given by Matz in [21] (see Lem. 3.2). In the same paper, the author wondered whether the condition was also sufficient, so that it would yield a characterization of  $\text{REC}$ . We here introduce a complexity function for picture languages that allows us to reformulate Lemma 3.2 as a necessary condition on the growth of the new complexity function, in analogy to Theorem 3.1. This new formulation will be used to negatively answer Matz's open question (see Sect. 6).

Let us introduce the following definitions.

**Definition 5.1.** Let  $A$  be a boolean square matrix of size  $k$ .  $A$  is a *fooling matrix* if there exists a permutation of its rows such that, in the resulting matrix  $B = \|b_{ij}\|$ , we have, for any  $1 \leq i \leq k$ ,  $b_{ii} = 1$  and, for any  $1 \leq i, j \leq k$  with  $i \neq j$ , if  $b_{ij} = 1$  then  $b_{ji} = 0$ .

**Definition 5.2.** Let  $L$  be a picture language. The *fooling complexity function*  $F_L(m)$  gives the size of the maximal fooling matrix that is a sub-matrix of  $M_{L(m)}$ .

**Remark 5.1.** Since every permutation matrix is a fooling matrix, for any picture language  $L$ , it holds  $P_L(m) \leq F_L(m)$ .

In analogy to Theorem 3.1, we can introduce a new necessary condition for the recognizability of picture languages, that rephrases Lemma 3.2 in the framework of Hankel matrices.

**Theorem 5.1.** *Let  $L \subseteq \Sigma^{**}$ . If  $L \in \text{REC}$  then there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $F_L(m) \leq c^m$ .*

*Proof.* For any  $m$ , consider a sub-matrix  $A_{(U,V)}$  of  $M_{L(m)}$  that is a fooling matrix. Let  $k$  be the size of  $A_{(U,V)}$  and let  $B = \|b_{u_i v_j}\|$  be the matrix obtained by a permutation of the rows of  $A_{(U,V)}$  such that, for any  $1 \leq i \leq k$ ,  $b_{u_i v_i} = 1$  and, for any  $1 \leq i, j \leq k$  with  $i \neq j$ , if  $b_{u_i v_j} = 1$  then  $b_{u_j v_i} = 0$ . Further, define  $P_m = \{(u_1, v_1), (u_2, v_2), \dots, (u_k, v_k)\} \subseteq \Sigma^{m,*} \times \Sigma^{m,*}$ . From the definition of a fooling matrix, it follows that  $P_m$  satisfies the condition of Lemma 3.2: for any  $1 \leq i, j \leq k$ ,  $b_{u_i v_i} = 1$  means that  $u_i \Phi v_i \in L$ , and the fact that for any  $i \neq j$ ,  $b_{u_i v_j}$  and  $b_{u_j v_i}$  are never simultaneously equal to 1, is equivalent to  $\{u_i \Phi v_j, u_j \Phi v_i\} \not\subseteq L$  for any  $i \neq j$ . Therefore,  $|P_m|$  is  $2^{O(m)}$ . Note that  $|P_m|$  is exactly the size of  $A_{(U,V)}$ . Hence, in particular,  $F_L(m)$ , the size of a maximal sub-matrix of  $M_{L(m)}$  that is a fooling matrix, is  $2^{O(m)}$ .  $\square$

### 6. CLASSES OF LANGUAGES DEFINED BY FUNCTIONS

In this section we consider a class of unary languages defined by functions, as introduced in [12], together with their stars and complements. When such languages are defined by super-exponential functions, we can obtain examples of languages in HP or HK, generalizing language  $L$  and its complement, as defined in Example 3.1. We will investigate the recognizability of the languages defined by functions. Then, using them as counter-examples, we will prove that the necessary conditions for recognizability stated in Theorem 3.1, Items 1 and 2, as well as the condition in Theorem 5.1, are not sufficient. This, in particular, answers the question stated by Matz in [21].

**Definition 6.1.** Given a function  $f$ , the picture language defined by  $f$ , is the set  $L_f = \{(m, f(m)) \mid m \geq 0\}$ .

We will also consider the languages  $L_f^{*\Phi}$ , that will be denoted  $L_{f^*}$ , and their complements  $\overline{L_f}$  and  $\overline{L_{f^*}}$ .

**Example 6.1.** Language  $L$  introduced in Example 3.1 is language  $\overline{L_{f^*}}$ , with  $f(m) = lcm(2^m + 1, \dots, 2^{m+1})$ .

Given a function  $f$ , if  $v$  denotes the picture with  $m$  rows and one column, the Hankel matrix associated to the language  $L_f(m)$ , carries a symbol 0 in all its positions except for positions indexed by  $(\varepsilon, v^{f(m)})$ ,  $(v, v^{f(m)-1})$ ,  $\dots$ ,  $(v^{f(m)-1}, v)$ ,  $(v^{f(m)}, \varepsilon)$ . In other words it is given by the following matrix “surrounded” by an infinite number of columns and rows of 0’s.

	$\varepsilon$	$v$	$v^2$	$\dots$	$v^{f(m)-1}$	$v^{f(m)}$	$v^{f(m)+1}$
$\varepsilon$	0	0	0	$\dots$	0	1	0
$v$	0	0	0	$\dots$	1	0	0
$v^2$	0	0	0	$\dots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v^{f(m)-1}$	0	1	0	$\dots$	0	0	0
$v^{f(m)}$	1	0	0	$\dots$	0	0	0
$v^{f(m)+1}$	0	0	0	$\dots$	0	0	0

The Hankel matrix for  $L_{f^*}(m)$  can be obtained by gluing an infinite number of copies of the following matrix.

	$v$	$v^2$	$v^3$	$\dots$	$v^{f(m)-2}$	$v^{f(m)-1}$	$v^{f(m)}$
$\varepsilon$	0	0	0	$\dots$	0	0	1
$v$	0	0	0	$\dots$	0	1	0
$v^2$	0	0	0	$\dots$	1	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v^{f(m)-3}$	0	0	1	$\dots$	0	0	0
$v^{f(m)-2}$	0	1	0	$\dots$	0	0	0
$v^{f(m)-1}$	1	0	0	$\dots$	0	0	0

By the analysis of the above Hankel matrices, we can state the following proposition.

**Proposition 6.1.** *Let  $f$  be a function. Then for any integer  $m$ :*

1.  $P_{L_f}(m) = F_{L_f}(m) = K_{L_f}(m) = f(m) + 1, R_{L_f}(m) = f(m) + 2.$
2.  $P_{\overline{L_f}}(m) = 2, F_{\overline{L_f}}(m) = 3, K_{\overline{L_f}}(m) = f(m) + 1, R_{\overline{L_f}}(m) = f(m) + 2.$
3.  $P_{L_{f^*}}(m) = F_{L_{f^*}}(m) = K_{L_{f^*}}(m) = R_{L_{f^*}}(m) = f(m).$
4.  $P_{\overline{L_{f^*}}}(m) = 2, F_{\overline{L_{f^*}}}(m) = K_{\overline{L_{f^*}}}(m) = R_{\overline{L_{f^*}}}(m) = f(m).$

*Proof.* Let us prove the equality  $F_{\overline{L_f}}(m) = 3$  of Item 2. The other statements can be easily proved. Recall that the Hankel matrix of the complement of a language can be simply obtained by exchanging 0 and 1 occurrences in the Hankel matrix for the language. Consider now, for any  $m$ , the Hankel matrix, say  $M$ , of the language  $\overline{L_f}(m)$  and remark that any row of  $M$  contains one symbol 0 at most. Let  $B$  be a fooling sub-matrix of  $M$ , of size  $k$  for some integer  $k$ , and let  $b_0$  the number of symbols 0 that occur in  $B$ . Obviously  $b_0 \leq k$ . The number of symbols 1 that occur in  $B$ , a part from the  $k$  on the counter-diagonal positions, is  $k^2 - k - b_0$ . Since  $B$  is fooling, it must be  $k^2 - k - b_0 \leq b_0$ , i.e. recalling  $b_0 \leq k, k^2 - k \leq 2k$  that is  $k \leq 3$ . Hence  $F_{\overline{L_f}}(m) \leq 3$ . But it is easy to find a fooling sub-matrix of  $M$  of size 3 and hence the equality  $F_{\overline{L_f}}(m) = 3$  follows.  $\square$

A first result on the recognizability of languages defined by functions is the following.

**Proposition 6.2** ([12]). *If  $L_f \in REC$  then  $f(m) \leq c^m$ .*

The proof of Proposition 6.2, is obtained in [12] by applying an Iteration Lemma [14]. Note that the proof can be also obtained from Theorem 5.1 remarking, as in Proposition 6.1, that  $F_{L_f}(m) = f(m) + 1$ .

From now on, let us consider super-exponential functions, i.e. functions  $f(m)$  for which there does not exist a constant  $c$  such that  $f(m) \leq c^m$ , for all  $m \geq 1$ . In this case, the associated two-dimensional languages  $L_f$  and their complements are not recognizable. To show this, let us first state a preliminary result.

**Lemma 6.1.** *Let  $\Sigma = \{a\}$  be a unary alphabet. Let  $w \in \Sigma^*$  and  $L$  be the string language  $L = \Sigma^* \setminus \{w\}$ . Then, every NFA recognizing  $L$  has at least  $|w| + 1$  states.*

*Proof.* Let  $\mathcal{A} = (Q, q_0, F, \delta)$  be a NFA recognizing  $L$ . Suppose that  $\mathcal{A}$  has less than  $|w| + 1$  states (i.e.  $|Q| < |w| + 1$ ) and consider all paths in  $\mathcal{A}$  labelled by  $w$  starting from  $q_0$ . Let us denote them  $\mathcal{P}_1, \dots, \mathcal{P}_k$  and let  $S$  be the set of the ending states of the paths  $\mathcal{P}_i$ , for  $i = 1, \dots, k$ . Since  $w \notin L$ , in  $S$  there are no final states. Moreover, since  $|Q| < |w| + 1$ , any  $\mathcal{P}_i$  contains at least a cycle, say  $\mathcal{C}_i$ , for  $i = 1, \dots, k$ , and, in any cycle, the state reachable from the initial state by a path of length  $|w|$  is not final. Let us denote by  $\lambda_i$  the length of  $\mathcal{C}_i$ , for  $i = 1, \dots, k$ , and let  $\lambda = lcm(\lambda_1, \dots, \lambda_k)$ . Consider now the word  $a^{|w|+\lambda}$ : the set of the ending states of paths in  $\mathcal{A}$  of length  $|w| + \lambda$  starting from  $q_0$  is exactly equal to  $S$ . But this is a contradiction since  $S$  contains no final state and  $a^{|w|+\lambda} \in L$ .  $\square$

**Proposition 6.3.** *Let  $f(m)$  be a super-exponential function. Then  $L_f, \overline{L_f} \notin REC$ .*

*Proof.*  $L_f \notin REC$  follows from Proposition 6.2.

Suppose now, by contradiction, that  $\overline{L_f} \in REC$ . From Lemma 3.1, there is  $c \in \mathbb{N}$  such that, for any  $m \geq 1$ , there exists an automaton, with  $c^m$  states at most, recognizing  $\overline{L_f}(m)$ . Recall that  $\overline{L_f}(m)$  can be viewed as a string language over the alphabet of the columns of height  $m$ , say  $\Sigma = \{v\}$  (see Lem. 3.1).

But this contradicts Lemma 6.1, since  $\overline{L_f}(m) = \Sigma^* \setminus \{v^{f(m)}\}$  and there does not exist a constant  $c$  such that  $f(m) \leq c^m$ , for all  $m \geq 1$ .  $\square$

**Example 6.2.** Let  $f(m) = m!$  and consider the language  $L_f = \{(m, m!) \mid m \geq 0\}$ . Since  $f(m) = m!$  is super-exponential, from Proposition 6.3, we have  $L_f, \overline{L_f} \notin REC$ .

Previous results and examples allow us to claim the following result. In particular this answers Matz’s open question in [21], whether the reverse of Lemma 3.2 holds, i.e. whether any language with “low” fooling complexity is recognizable.

Indeed, there is a one-to-one correspondence between a fooling matrix in  $M_{L(m)}$  and a set  $\{P_m\}$  as considered in Lemma 3.2 since, for all pairs  $(p, q), (p', q')$  in  $P_m$ , we have  $p \neq p'$  and  $q \neq q'$ .

**Proposition 6.4.** *The necessary conditions for recognizability stated in Theorem 5.1 and in Theorem 3.1, Items 1 and 2, are not sufficient.*

*Proof.* Examples of languages not in REC with both permutation complexity and fooling complexity less than exponential are all languages  $\overline{L_f}$  where  $f(m)$  is a super-exponential function. Indeed (see Prop. 6.1), given a function  $f(m)$ , the language  $\overline{L_f}$  is such that  $P_{\overline{L_f}}(m) = 2$  and  $F_{\overline{L_f}}(m) = 3$ . Moreover, if  $f(m)$  is a super-exponential function, then from Proposition 6.3,  $\overline{L_f} \notin REC$ . This implies that the necessary conditions for languages in REC, given in Theorem 5.1 and in Item 2 of Theorem 3.1, are not sufficient.

Languages  $L_f$  where  $f(m)$  is a super-exponential function and  $f(m) \leq 2^{c^m}$ , for some constant  $c$ , are examples of languages not in  $REC \cup \text{co-REC}$  with row

complexity less than  $2^{c^m}$ . Indeed (see Prop. 6.1)  $R_{L_f}(m) = f(m) + 2 \leq 2^{d^m}$ , while  $L_f, \overline{L_f} \notin \text{REC}$  as in Proposition 6.3.

In particular language  $\overline{L_f}$  with  $f(m) = m!$  (see Ex. 6.2) is an example of non-sufficiency of all three conditions, since  $m! \leq m^m \leq 2^{c^m}$ , for some  $c$  and sufficiently large  $m$ .  $\square$

The situation for languages  $L_{f^*}$  and their complements is different. Even if  $f(m)$  is super-exponential,  $\overline{L_{f^*}}$  can belong to REC (consider language  $L$  in Exs. 3.1 and 6.1).

**Proposition 6.5.** *If  $f(m)$  is a super-exponential function, then  $L_{f^*} \notin \text{REC}$ . Moreover, if  $\overline{L_{f^*}} \in \text{REC}$ , then  $\overline{L_{f^*}} \notin \text{UREC}$  and  $\overline{L_{f^*}}$  is infinitely ambiguous.*

*Proof.*  $L_{f^*} \notin \text{REC}$  follows from Theorem 5.1, recalling that  $F_{L_{f^*}}(m) = f(m)$  (see Prop. 6.1). Further if  $\overline{L_{f^*}} \in \text{REC}$ , then  $L_{f^*} \in \text{HP}$  and  $\overline{L_{f^*}} \in \text{HK}$  by Corollary 4.1. Hence  $\overline{L_{f^*}} \notin \text{UREC}$ , from Proposition 4.2, while  $\overline{L_{f^*}}$  is infinitely ambiguous, from Proposition 4.3.  $\square$

## 7. CONCLUSIONS AND OPEN QUESTIONS

In the paper we afforded some open problems on unambiguity, finite ambiguity and complementation (Questions 1, and 2 in the introduction) and gave some partial answers; furthermore we dealt with some recognizability conditions for picture languages and showed that they are not sufficient.

A complete answer to Question 1 seems far to be found, also due to its interpretation in the computational complexity framework. With regards to Question 2, we considered the possibility that in REC there exist finitely ambiguous languages, and showed that this is not true for a class of languages in REC. Note that in a bit different framework (see [5]), when the recognition is accomplished without border symbols (tiles with  $\#$  are not allowed), it is shown that there is an infinite hierarchy of finitely ambiguous languages. Therefore the border symbols have to play a major role, in order to show that in REC there do not exist finitely ambiguous languages. We figure that when a language is recognized by a tiling system with finite ambiguity, then it is possible to obtain an unambiguous tiling system, for the language, by paying special attention to border tiles.

For what concerns the Matz's condition for languages in REC (Lem. 3.2 or equivalently Thm. 5.1), remark that the languages we considered as examples of non-sufficiency of this condition are neither in REC nor in co-REC. We conjecture a new formulation of the recognizability condition for two-dimensional languages: "Let  $L \in \text{co-REC}$ ;  $L \in \text{REC}$  if and only if there is a  $c \in \mathbb{N}$  such that, for all  $m \geq 1$ ,  $F_L(m) \leq c^m$ ".



## REFERENCES

- [1] M. Anselmo and M. Madonia, Deterministic and unambiguous two-dimensional languages over one-letter alphabet. *Theoret. Comput. Sci.* **410** (2009) 1477–1485.
- [2] M. Anselmo and M. Madonia, A note on unambiguity, finite ambiguity and complementation in recognizable two-dimensional languages, in *Proc. CAI 09. Lect. Notes Comput. Sci.* **5725** (2009) 147–159.
- [3] M. Anselmo, D. Giammarresi, M. Madonia and A. Restivo. Unambiguous recognizable two-dimensional languages. *RAIRO-Theor. Inf. Appl.* **40** (2006) 227–294.
- [4] M. Anselmo, D. Giammarresi and M. Madonia, From determinism to non-determinism in recognizable two-dimensional languages, in *Proc. DLT 07. Lect. Notes Comput. Sci.* **4588** (2007) 36–47.
- [5] M. Anselmo, N. Jonoska and M. Madonia, Framed versus unframed two-dimensional languages, in *Proc. SOFSEM 09. Lecture Notes in Comput. Sci.* **5404** (2009) 79–92.
- [6] M. Anselmo, D. Giammarresi and M. Madonia, Deterministic and unambiguous families within recognizable two-dimensional languages. *Fund. Inform.* **98** (2010) 143–166.
- [7] A. Bertoni, M. Goldwurm and V. Lonati, The complexity of unary tiling-recognizable picture languages. *Fund. Inform.* **90** (2009) 231–249.
- [8] J.-C. Birget, Intersection and union of regular languages and state complexity. *Inform. Process. Lett.* **43** (1992) 185–190.
- [9] S. Brocchi, *Bidimensional pictures: reconstruction, expression and encoding*, Ph.D. thesis. <http://www.dsi.unifi.it/DRIIA/RaccoltaTesi/Brocchi.pdf>
- [10] J. Cervelle, *Langages de figures*, Rapport de stage, ENS Lyon (1997).
- [11] S. Eilenberg, *Automata, Languages and Machines*, Vol. A. Academic Press (1974).
- [12] D. Giammarresi, Two-dimensional languages and recognizable functions, in *Proc. DLT 93*, edited by G. Rozenberg and A. Salomaa. World Scientific Publishing Co., Singapore (1994), 290–301.
- [13] D. Giammarresi and A. Restivo, Recognizable picture languages. *Int. J. Pattern Recogn. Artif. Intell.* **6** (1992) 241–256.
- [14] D. Giammarresi and A. Restivo, *Two-dimensional languages, Handbook of Formal Languages*, Vol. III. G. Rozenberg et al., Eds. Springer Verlag (1997), 215–268.
- [15] D. Giammarresi and A. Restivo, Matrix based complexity functions and recognizable picture languages, in *Logic and Automata: History and Perspectives*, E. Grader, J. Flum and T. Wilke, Eds. Texts in Logic and Games 2. Amsterdam University Press (2007), 315–337.
- [16] D. Giammarresi and A. Restivo, Ambiguity and complementation in recognizable two-dimensional languages, in *Proc. Int. Conf. Theoret. Comput. Sci., IFIP*, Vol. **273**, edited by G. Ausiello, J. Karhumäki, G. Mauri and L. Ong. Springer, Boston (2008), 5–20.
- [17] I. Glaister, J. Shallit, A lower bound technique for the size of nondeterministic finite automata. *Inform. Process. Lett.* **59** (1996) 75–77.
- [18] J. Hromkovic, *Communication Complexity and Parallel Computing*. Springer (1997).
- [19] J. Hromkovic, J. Karumäki, H. Klauck, G. Schnitger and S. Seibert, Communication complexity method for measuring nondeterminism in finite automata. *Inform. Comput.* **172** (2002) 202–217.
- [20] V. Lonati and M. Pradella, Snake-deterministic tiling systems, in *Proc. MFCS 2009, 34th International Symposium on Mathematical Foundations of Computer Science. Lect. Notes Comput. Sci.* **5734** (2009) 549–560.
- [21] O. Matz, On piecewise testable, starfree, and recognizable picture languages, in *Foundations of Software Science and Computation Structures*, Vol. **1378**, M. Nivat, Ed. Springer-Verlag, Berlin (1998).
- [22] O. Matz, *Dot-depth and monadic quantifier alternation over pictures*, Ph.D. thesis Technical Report 99-08, RWTH Aachen (1999).

- [23] O. Matz, Dot-depth, monadic quantifier alternation, and first-order closure over grids and pictures, *Theoret. Comput. Sci.* **270** (2002) 1–70.
- [24] I. Mäurer, *Characterizations of Recognizable Picture Series*, Ph.D. thesis, Universität Leipzig, Institut für Informatik, Abteilung Automaten und Sprachen (2007).
- [25] A. Pothhoff, S. Seibert and W. Thomas, Nondeterminism versus determinism of finite automata over directed acyclic graphs. *Bull. Belgian Math. Soc.* **1** (1994) 285–298.
- [26] K. Reinhardt, The  $\#a = \#b$  Pictures are recognizable, in *Proc. 18th STACS 2001. Lect. Notes Comput. Sci.* **2010** (2001) 527–538.

Communicated by A. Restivo.

Received February 27, 2010. Accepted November 4, 2010.